

主体と動詞の属性に基づく複文の接続構造の解析

向 伸 景頁 †

本論文は、動詞と主体の属性を用いて複文の接続関係の关系的意味を解析し、この関係的意味を用いて接続構造を解析するモデルについて述べる。従来、複文の従属節間の接続構造解析は、接続の表現を階層的に分類し、その階層的な順序関係による方法が用いられてきた。しかし、接続の表現には曖昧性があり、同じ接続の表現でも意味が違ふと係り方が違ふ。このため、本論文では、動詞と主体の属性を用いて、接続関係の意味を求め、この接続関係の关系的意味を、接続関係の距離によって分類する。この分類を用いて接続構造を解析する方法を用いた。動詞の属性として、意志性、アスペクト・ムード・ヴォイス、意味分類などを用いた。主体の属性として、主節と従属節の主体が同一かどうか、無生物主体かどうかを採用した。このモデルを実際の用例により評価した結果 98.4%の精度が得られた。接続の表現の階層的分類を使用したモデルに同じ用例を適用したところ 97.0%の精度が得られたので、本論文のモデルを使用することにより誤り率が約半分に改善された。

キーワード: 接続構造, 複文, 動詞の属性, 主体の属性

A Model for Analyzing Structures of Coherence Relations Using Features of Verbs and Subjects

KOU MUKAINAKA †

The model proposed in this paper analyzes meanings of coherence relations using features of verbs and subjects in Japanese complex sentences, and then analyzes structures of coherence relations using the meanings of coherence relation. Dependency structures of Japanese complex sentences are usually analyzed using hierarchical classification of conjunctions and conjunctive particles. But, conjunctions and conjunctive particles usually have multiple senses and are ambiguous. If a conjunction or conjunctive particle in a subordinate clause has a different sense, the subordinate clause may modify a predicate in a different clause. So, the model analyzes the coherence relations between subordinate clauses and a main clause using the features of verbs and subjects, and defines the meanings of coherence relations. Then the meanings of coherence relations are classified according to distance of coherence. The model uses this classification of coherence relations to analyze the structures of coherence relations. Volition, aspect, mood, voice and semantic category etc. are used as the features of verbs, and animate or inanimate etc. is used as the features of subjects. The model is evaluated by examples from actual documents, and shows 98.4% accuracy. Since the model using the classification of conjunctions and conjunctive particles shows 97.0% accuracy with the same examples, the model proposed in this paper, decreases the error rate by half.

KeyWords: *Coherence structure, Complex sentences, Attributes of verbs, Attributes of subjects*

† 江戸川大学社会学部 環境情報学科, Department of Environmental Information, Edogawa University

1 はしがき

日本語の複文の従属節には、体言に係る連体修飾節と、用言に係る連用修飾節がある。連体修飾節は通常次の句の体言に係る場合が多く曖昧性は比較的少ない。ところが、連用修飾節は係り先に曖昧性があり、必ずしもすぐ次の節の用言に係るとは限らない。

このような曖昧性を解消するために、接続助詞、接続詞など接続の表現を階層的に分類し、その順序関係により、接続関係を解析する方法(白井、池原 1995)が用いられてきた。また、接続関係を、接続の表現を基に統計的に分析し、頻度の高い接続関係を優先する方法(宇津呂、西岡山 1999)も用いられてきた。

しかし、接続の表現には曖昧性があり、同じ接続の表現でも異なる意味で用いられるときは異なる係り方をする。従って、接続の表現の階層的な分類を手がかりとする方法では、達成できる精度に限界がある。

本論文では、従属節の動詞と主体の属性を用いて接続関係の关系的意味を解析し、接続構造を解析する方法を用いる。本方法によりモデルを作成し、解析した結果と従来から行われてきた接続の表現の表層的な分類を用いた方法とを同じ例文を用いて比較する。ここで、主体は「複文の研究」(仁田 1995)で使っているのと同じ意味で使っており、後述の解析モデルでは「が格」として処理している。

2 動詞と主体の属性と接続関係の关系的意味

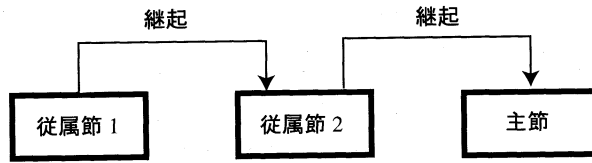
接続助詞、接続詞などの接続の表現には曖昧性がある。特に、「て」、「が」、「と」などであらわされる従属節の关系的意味はさまざまである。これは「から」、「ので」などと違って、これらの接続の表現自体が明確な固定的意味を有していないからである。従って、これらの接続の表現では、接続関係の关系的意味が、従属節や主節の表している事象の意味、およびそれらの事象の相互関係によって決まってくる。

事象の意味は、ある主体が行う行動または状態の分類、意志性などをあらわす。たとえば、生物主体の姿勢変化、心的状態などである。従って、事象の意味は、動詞と名詞の属性を用いて表すことができると考えることができる。

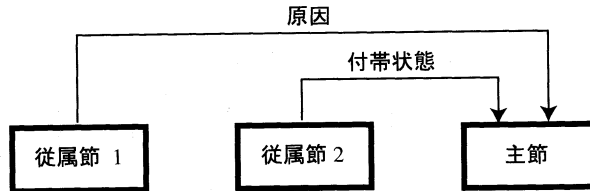
たとえば、助詞「て」による接続関係で、従属節と主節が共に意志動詞で形成され、両者が同一主体のときは、「時間的継起」を表すことが多い。

【例文1】 ユーザは、ID とパスワードを指定して、OK ボタンをクリックします。

同じ助詞「て」による接続関係でも、従属節の動詞が、姿勢変化、携帯・保持、心的状態などの意味分類であるときは「付帯状態」を表すことが多い。「付帯状態」とは、従属節と主節の事象が同じ時間の中に存し、同一主体で、従属節で主節の事象の主たる動きや状態の実現され方を限定・修飾するものである。



(a) 例文4の接続関係



(b) 例文5の接続関係

図 1 接続関係の意味による係り方の違い

〔例文 2〕 そこで女中が鍵を持って、私を待っていた。

「て」による接続関係で、従属節と主節が共に無意志動詞で、従属節が無生物主体、主節が生物主体のときは、「原因」を表すことが多い。これは、主たる事象が人間を表すものでありながら、従属節で人間の意志で制御できない事象が生じたためである。

〔例文 3〕 彼は、車が故障して、遅れた。

このような、接続関係に影響を与える動詞の属性としては、意志性、意味分類の外に、ヴォイス、アスペクト、ムード、慣用句の分類、主節と従属節動詞の類似性などがある。主体の属性としては、生物／無生物性、主節と従属節の主体の類似性を用いる。

3 接続の表現の意味による距離

従属節の係り先は、一般的には次の従属節であることが多いが、さらに先の従属節に係ることもある。例えば、次の例文を考えてみよう。

〔例文 4〕 クエリーボタンを押して、メニューから実行を選択して、検索情報を表示する。

〔例文 5〕 彼は、連絡がないかと思って、携帯電話を持って、待っていた。

〔例文 4〕では図 1 (a) に示すように、隣接する従属節に係っているが、〔例文 5〕では隣接する従属節を飛び越えて主節に係っている。このように、同じ助詞「て」による接続関係でも、接続関係の意味の違いによって係り方が異なってくる。

一般的に、接続関係の意味の違いによって、接続関係の距離に違いがあり、係る文の述語に「密着している」と「離れている」ものがある。「密着している」接続関係は、隣接する従属節に係りやすく、「離れている」接続関係はより遠くへ係りやすい。

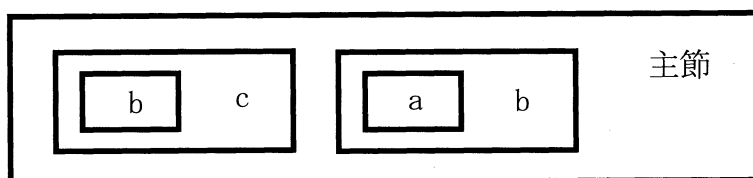


図 2 連用修飾節の包含関係

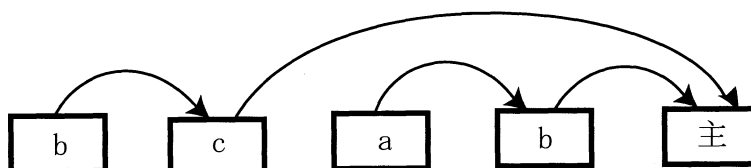


図 3 連用修飾節の係り受け

表 1 接続関係の距離

接続の表現のクラス	距離
A	1
B	2
C	3

3.1 接続の表現を A, B, C の 3 分類にして解析した場合

南 (南 1993) は接続助詞や連用中止などの接続の表現による連用修飾節を A, B, C, の 3 クラスに分類して, 図 2 の例に示すように, A は A のみを含むことができ, B は A, B を含むことができ, C は A, B, C を含むことができることを示した. 主節は A, B, C 共に含むことができる.

これに, 他の連用修飾節を飛び越えることができるかどうかという関係を考慮に入れて, 係り受けの関係で表すと図 3 のようになる. A は A, B, C に係ることができ, B は A を飛び越えて B, C に係ることができ, C は A, B を飛び越えて C に係ることができると表現することができる. すなわち, A は「密着している」接続関係であり, 隣接する従属節に係りやすく, C は「離れている」接続関係であって, より遠くへ係りやすい. B はその中間である. これを表 1 に示すように, 接続関係の距離で表す.

連用修飾節の距離を m , 係り先の節の距離を d , 飛び越える節の距離を j で表したとき, 接続関係のルールは次の 2 つの条件で表すことができる.

〔ルール 1〕 連用修飾節は自身の距離以上の距離を持つ節に係る. ($d \geq m$)

〔ルール 2〕 連用修飾節は自身の距離より大きな距離の節を飛び越えて係ることはできない. ($j < m$)

表 2 用例における係り受け接続関係の頻度
(A, B, C の 3 分類で接続の表現の頻度による)

係り側	受け側			
	A	B	C	主節
A	0	21	5	28
B	3	197	46	414
C	0	5	20	132

表 3 用例における接続関係で係り側と飛び越える節の頻度
(A, B, C の 3 分類で接続の表現の頻度による)

係り側	飛び越える節の距離が最大の節			
	隣接	A	B	C
A	52	0	2	0
B	508	22	122	8
C	58	5	91	3

このルールが実際の用例について、どの程度よく適合するかを調べた。複文に関する論文集(仁田 1995)、日本語教育の参考書(北条 1992)、ネットワークの解説書(金内, 今安 1993)の 2,010 文から、主節の外に連用修飾節を 2 つ以上含む 344 文を取り出して用例とした。

表 2 は係り側と受け側の接続関係の頻度を調べて集計したものである。

この表に前記のルール 1 を適用すると、右上から対角線のセルまでがこのルールで正しく処理される接続関係であり、それより左下のセルは、このルールでは間違って処理される接続関係である。

飛び越えて係る場合の連用修飾節と飛び越える節の距離が最大の節(最も離れているもの)との関係について頻度を調べて集計した結果を表 3 に示す。

この表に、ルール 2 を適用すると左下から対角線までが、このルールで正しく処理される接続関係であり、それより右上のセルはこのルールでは間違って処理される接続関係である。

南の A, B, C の分類で、いくつかの接続の表現は、複数の分類に含まれる。例えば、接続助詞「て」による接続関係は、接続関係の意味の違いによって A にも B にも含まれる。表 2, 表 3 の集計は、接続の表現だけを見て分類集計したものであって、同じ接続の表現が複数に分類されるときは、最も頻度の高い分類を採用した。

前節で述べた接続関係の関係の意味を調べて、関係の意味によって分類し、係り受けを解析したほうが、より正確な解析結果が出るはずである。このような観点から、接続の表現を接続関係の関係の意味によって、表 4 に示すように A, B, C に 3 分類した。

表 4 接続関係の意味分類 (A, B, C の 3 分類)

意味分類	接続関係の関係の意味	接続の表現の例
A	同時動作, 方法, 手段, 携帯, 心的状態	て (付帯状態), て (方法), ながら (同時動作)
B	継起, 条件, 時間, 程度, 原因, 理由, 目的	て (継起), し (継起), 連用中止 (継起), せず (継起), ないで (継起), と (条件), と (時), ば (条件), たら (条件), たら (時), ても (逆接条件), ばあい (時), さいに (時), さい (時), ころ (時), のに (逆接条件), ほど (程度), て (原因), て (理由), し (理由), から (原因), から (理由), ので (原因), ので (理由), ため (目的), ため (理由), たら (理由), よう (目的), ように (目的), ことで (理由)
C	前提, 前置き, 逆接	が (逆接), が (前提), ように (前提), ながら (逆接)

表 5 用例における係り受け接続関係の頻度 (A, B, C の 3 分類で接続関係の関係の意味による)

係り側	受け側			
	A	B	C	主節
A	4	39	13	62
B	1	160	57	372
C	0	3	20	140

表 6 用例における接続関係で係り側と飛び越える節の頻度 (A, B, C の 3 分類で接続関係の関係の意味による)

係り側	飛び越える節の距離が最大の節			
	隣接	A	B	C
A	117	0	0	1
B	461	34	90	5
C	57	12	92	2

表 4 に基づいて前記と同じ 344 文の用例を分析して集計した結果が表 5, 表 6 である。

これらの表から接続の表現の頻度による場合と, 接続関係の関係の意味を分析した場合について, 精度を計算すると表 7 に示すようになる。

この表から分るとおり, 主節以外に連用修飾節を 2 つ以上含む 344 文中の個々の接続関係を

表 7 A, B, C の 3 分類を用いた場合の接続構造の解析精度 (単位: %)

接続関係の精度か文単位での精度かの区別	接続の表現 の頻度によ る解析結果	接続関係の 関係の意味 による解析 結果
接続関係の精度 (344 文中の 871 の接続関係)	83.5	88.2
主節のほかに連用従属節を 2 つ以上含む文での精度 (344 文)	62.2	72.1
全体の文での精度 (2010 文)	93.5	95.7

集計した結果では、正しく解析できなかった接続関係が 16.5%から 11.8%に改善した。これを、344 文中で正しく解析できなかった文の比率で見ると、37.8%から 27.9%に、全体の 2,010 文で正しく解析できなかった文の比率では、6.5%から 4.3%に改善した。

3.2 接続の表現を 5 分類した場合

前期の A, B, C の 3 分類で、A, C に比べて B に含まれる接続の表現が非常に多い。これが、解析の精度が上がらない原因になっている。このため、接続関係の関係の意味により、B を継起, 条件, 原因に 3 分類する。もともと、A の接続関係の関係の意味は付帯状態を表し、B は前提を表しているの、接続関係の意味分類を整理すると表 8 に示すようになる。

表で、並列節だけは別分類にして表示してあるが、これは、並列節だけ異なったルールが適用されるためである。並列のスコープ内で、条件節などが並列節に係って並列の要素を構成するときは、並列節は前提と同じ距離で処理する。並列節が並列を構成する次の節に係るときは、並列のスコープ内で、並列節は付帯状態と同じ距離で処理する。

表 8 の接続関係の意味分類に基づき接続の表現を 5 分類し、係り側と受け側の接続関係の頻度および係り側と飛び越える節の頻度を調べて集計した。同じ接続の表現で接続関係の関係の意味の違いにより、複数の分類に含まれる接続の表現は、最も頻度の高い分類を適用した。並列節のスコープの決定は、並列を表す接続の表現を有する節で最も狭いスコープを採用した。解析の結果を表 9, 表 10 に示す。

接続関係の関係の意味を調べて、関係の意味により接続関係を意味分類し、係り側と受け側の頻度および飛び越える節の頻度を調べて集計した。並列節のスコープの決定は、並列を意味する接続の表現および並列節を構成する各々の節の述語, 目的語などの類似度によった。すなわち、接続の表現が並列を意味し、並列を構成する各々の節の述語, 目的語, 主語などが上位の意味分類で同じであれば、並列を構成するものとした。解析の結果を表 11, 表 12 に示す。

これらの表から、接続の表現の頻度による場合と、接続関係の関係の意味を分析した場合について、精度を計算すると表 13 に示すようになる。

表 8 接続関係の意味分類

意味分類	接続関係の关系的意味	接続の表現の例
付帯状態	同時動作, 方法, 手段, 携帯, 心的状態	て (付帯状態), て (方法), ながら (同時動作)
継起	継起	て (継起), し (継起), 連用中止 (継起), せず (継起), ないで (継起)
条件	条件, 時間, 程度	と (条件), と (時), ば (条件), たら (条件), たら (時), ても (逆接条件), ばあい (時), さいに (時), さい (時), ころ (時), のに (逆接条件), ほど (程度)
原因	原因, 理由, 目的	て (原因), て (理由), し (理由), から (原因), から (理由), ので (原因), ので (理由), ため (目的), ため (理由), たら (理由), よう (目的), ように (目的), ことで (理由)
前提	前提, 前置き, 逆接	が (逆接), が (前提), ように (前提), ながら (逆接)
並列	並列, 対比	て (並列), し (並列), が (対比), 連用中止 (並列), ば (並列), たり (並列), せず (並列), ように (対比), より (対比)

表 9 用例における接続関係で係り側と受け側の頻度 (5 分類で接続の表現の頻度による)

係り側	受け側					
	1 付帯状態	2 継起	3 条件	4 原因	5 前提	主節
1. 付帯状態	0	9	8	4	5	28
2. 継起	3	51	26	32	22	162
3. 条件	0	21	26	30	13	138
4. 原因	0	5	0	6	11	114
5. 前提	0	1	2	2	20	132

この表から, 5 分類を採用した場合は, 正しく解析できなかった接続関係の比率が, 接続の表現の頻度による場合は, 16.5%から 11.0%に, 接続関係の关系的意味による場合は, 11.8%から 4.8%に改善されたことが分る. 接続関係の关系的意味による場合は 2 倍以上改善される.

表 10 用例における接続関係で係り側と飛び越える節の頻度（5 分類で接続の表現の頻度による）

係り側	飛び越える節の距離が最大の節					
	隣接	1 付帯状態	2 継起	3 条件	4 原因	5 前提
1. 付帯状態	52	0	2	0	0	0
2. 継起	246	9	16	15	8	2
3. 条件	165	7	45	7	0	4
4. 原因	97	6	13	16	2	2
5. 前提	58	5	26	36	29	3

表 11 用例における接続関係で係り側と受け側の頻度（5 分類で接続関係の关系的意味による）

係り側	受け側					
	1 付帯状態	2 継起	3 条件	4 原因	5 前提	主節
1. 付帯状態	4	11	11	17	13	62
2. 継起	1	39	23	20	18	117
3. 条件	0	7	27	32	23	132
4. 原因	0	2	0	10	16	123
5. 前提	0	1	1	1	20	140

表 12 用例における接続関係で係り側と飛び越える節の頻度（5 分類で接続関係の关系的意味による）

係り側	飛び越える節の距離が最大の節					
	隣接	1 付帯状態	2 継起	3 条件	4 原因	5 前提
1. 付帯状態	117	0	0	0	0	1
2. 継起	193	10	3	7	4	1
3. 条件	162	12	39	5	0	3
4. 原因	106	12	16	15	1	1
5. 前提	57	12	20	40	32	2

表 13 5 分類を用いた場合の接続構造の解析精度（単位：％）

接続関係の精度か文単位の精度かの区別	接続の表現の頻度による解析結果	接続関係の关系的意味による解析結果
接続関係の精度（344 文中の 871 の接続関係）	89.0	95.2
主節のほかに連用従属節を 2 つ以上含む文での精度（344 文）	78.8	98.5
全体の文での精度（2010 文）	96.4	98.2

表 14 8 分類の接続関係の意味分類

意味分類	距離
付帯状態	1
継起	2
条件	3
原因	4
継起+カンマ	5
条件+カンマ	6
原因+カンマ	7
前提	8

表 15 用例における接続関係で係り側と受け側の頻度（8 分類で接続の表現の頻度による）

係り側	受け側								
	1 付 帯状 態	2 継 起	3 条 件	4 原 因	5 継 起, 件,	6 条 件, 因,	7 原 因, 提	8 前 提	主節
1. 付帯状態	0	0	0	1	9	8	3	5	28
2. 継起	0	2	1	0	4	7	3	1	13
3. 条件	0	0	1	1	4	1	2	2	18
4. 原因	0	0	0	0	2	0	0	2	10
5. 継起 + カンマ	3	0	0	2	45	18	27	21	149
6. 条件 + カンマ	0	0	0	1	17	24	26	11	120
7. 原因 + カンマ	0	0	0	0	3	0	6	9	104
8. 前提	0	0	1	0	1	1	2	20	132

3.3 接続の表現を 8 分類した場合

一般的に接続関係で、連用修飾節にカンマ（または読点）のある場合とない場合で、従属節間の距離に違いが出る。カンマのない節の方がより「密着している」接続関係であり、カンマのある節の方がより「離れている」接続関係であるといえる。

前節の 5 分類の解析結果を見ると、継起、条件、原因の意味分類で誤りが多く出ていることが分る。さらに個々のケースを分析すると、個々の意味分類の差より、カンマが付くか付かないかの差の方が大きいことが分る。このため、表 14 に示すように意味分類と距離を定義する。

表 14 を用いて、接続の表現の頻度による場合と、接続関係の关系的意味による場合について、解析した結果を表 15、表 16、表 17、表 18 に示す。

表 16 用例における接続関係で係り側と飛び越える節の頻度（8 分類で接続の表現の頻度による）

係り側	飛び越える節の距離が最大の節								
	隣接	1 付 帯状 態	2 継 起	3 条 件	4 原 因	5 継 起, 件,	6 条 件, 因,	7 原 因, 提	8 前 提
1. 付帯状態	52	0	0	0	0	2	0	0	0
2. 継起	31	0	0	0	0	0	0	0	0
3. 条件	26	0	2	0	0	1	0	0	0
4. 原因	14	0	0	0	0	0	0	0	0
5. 継起 + カンマ	215	9	5	5	5	11	10	3	2
6. 条件 + カンマ	139	7	2	1	0	40	6	0	4
7. 原因 + カンマ	83	6	1	3	1	12	13	1	2
8. 前提	58	5	3	4	2	23	32	27	3

表 17 用例における接続関係で係り側と受け側の頻度（8 分類で接続関係の关系的意味による）

係り側	受け側								
	1 付 帯状 態	2 継 起	3 条 件	4 原 因	5 継 起, 件,	6 条 件, 因,	7 原 因, 提	8 前 提	主節
1. 付帯状態	4	1	2	2	10	9	15	13	62
2. 継起	0	1	0	0	3	6	2	1	12
3. 条件	0	0	1	1	2	2	2	4	18
4. 原因	0	0	0	0	1	0	0	3	10
5. 継起 + カンマ	1	0	0	1	35	17	17	17	105
6. 条件 + カンマ	0	0	0	1	5	24	28	19	114
7. 原因 + カンマ	0	0	0	0	1	0	10	13	113
8. 前提	0	0	0	0	1	1	1	20	140

これらの表から、接続の表現の頻度による場合と、接続関係の关系的意味を解析した場合について、精度を計算すると表 19 に示すようになる。

表 19 から、意味解析を伴わない場合は 5 分類の 11.0% から 8.8% に、意味解析を行った場合は 5 分類の 4.8% から 3.2% に改善されることが分る。意味解析を伴った場合の改善効果が大きい。

表 18 用例における接続関係で係り側と飛び越える節の頻度（8 分類で接続関係の关系的意味による）

係り側	飛び越える節の距離が最大の節								
	隣接	1 付 带状 態	2 継 起	3 条 件	4 原 因	5 継 起,	6 条 件,	7 原 因,	8 前 提
1. 付带状態	117	0	0	0	0	0	0	0	1
2. 継起	25	0	0	0	0	0	0	0	0
3. 条件	27	1	2	0	0	0	0	0	0
4. 原因	14	0	0	0	0	0	0	0	0
5. 継起 + カンマ	168	10	2	4	3	1	3	1	1
6. 条件 + カンマ	135	11	2	3	0	35	2	0	3
7. 原因 + カンマ	92	12	2	3	1	14	12	0	1
8. 前提	57	12	3	4	3	17	36	29	2

表 19 8 分類を用いた場合の接続構造の解析精度（単位：％）

接続関係の精度か文単位での精度かの区別	接続の表現 の頻度によ る解析結果	接続関係の 关系的意味 による解析 結果
接続関係の精度（344 文中の 871 の接続関係）	91.2	96.8
主節のほかに連用従属節を 2 つ以上含む文での精度（344 文）	82.3	93.3
全体の文での精度（2010 文）	97.0	98.9

4 接続関係の意味による接続構造の解析モデル

接続構造の解析モデルは図 4 に示すように、まず、動詞と主体の属性を用いて接続関係の意味解析を行い、その結果に基づいて並列節の解析、接続構造の解析を行う。

4.1 動詞と主体の属性を用いた接続関係の意味解析

接続関係の意味解析を行うときに動詞と主体の属性を用いるが、これらは表 20 に示す素性として表される。これらの素性のうちで、動詞の意志性は IPAL 辞書（情報処理振興事業協会 1987）のものをを用いた。意味分類は分類語彙表（国立国語研究所 1989）の分類を用いた。

これらの属性を用いて動詞の格パターン、接続助詞の接続関係パターンを HPSG の素性構造に似た形式で表し、辞書に登録した。

「疲れが出る」に対応する動詞「出る」の格パターンと後置詞句「疲れが」、名詞句「疲れ」

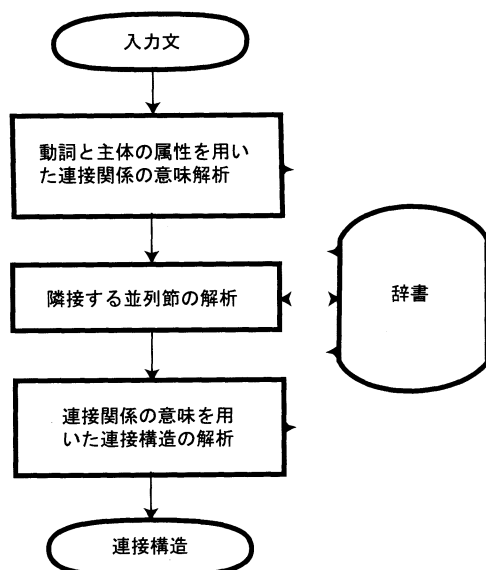


図 4 接続構造の解析モデル

表 20 動詞と主体の素性

素性	値	意味
VOLITION	+, -	意志性
ANIMATE	+, -	生物／無生物
SEM-CAT	使用, 製造, 教育, ...	意味分類
VOICE	受動, 能動, 使役, 可能	ヴォイス (受動／能動／使役／可能)
ASPECT	ている, である, ...	アスペクト
MODE	平叙, 疑問, 命令, 仮定, 指示	モード (平叙／疑問／命令／仮定／指示)
IDIOM-CAT	TE-SEQ, GA-PRE, ...	慣用句の分類

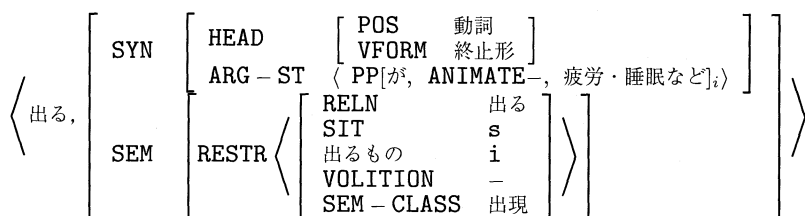
の記載例を図 5 に示す。

接続助詞「て」が「原因」を表す場合と、「付帯状態」を表す場合の接続関係パターンの記載例を図 6 に示す。

「原因」を表す接続関係は、従属節と主節が共に無意志動詞で、従属節が無生物主体、主節が生物主体を表す場合に対応する。「付帯状態」を表す接続関係は、従属節の動詞が「包摂、姿勢変化、着脱、携帯、心的変化など」の意味分類である場合に対応する。

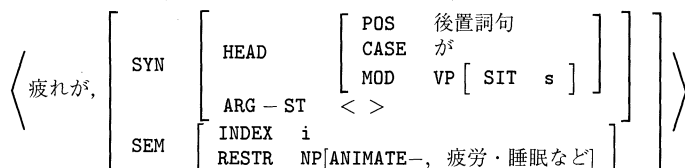
接続関係パターンは、表 20 の素性を用いて意味解析するもので、76 の接続関係パターン (向仲 1997) を用意した。図 5、図 6 では HPSG の形式で記載しているが、実際の辞書では prolog

(a) 動詞の記載例



注) VP[終止形, VOLITION-, ANIMATE-, 出現]。と略記する。

(b) 後置詞句: PP[が, ANIMATE-, 疲労・睡眠など] の内容



(c) 名詞:NP[ANIMATE-, 疲労・睡眠など] の内容

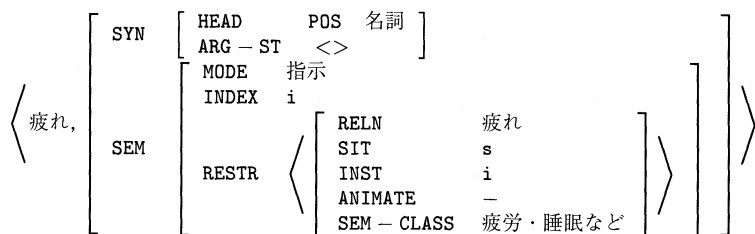


図 5 辞書における動詞の格パターンの記載例

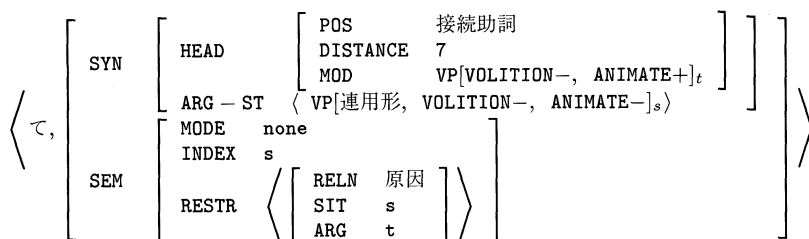
の形式に変換して格納している。

これら辞書の情報を用いて、接続関係の意味解析を行い、接続構造の解析を行って、複文の接続構造を生成する。解析は、簡単な HPSG パーザを prolog で作成して行った。このパーザは、主として格パターンの解析と接続関係パターンの解析の機能だけを持ったもので、この目的のために作成した。本論文のモデルは連用修飾節の解析を対象としたもので、一部の連体修飾節は入力文から省いてある。

連用修飾節「疲れが出て」に格パターンと接続関係パターンを適用して解析した例を図 7 に示す。

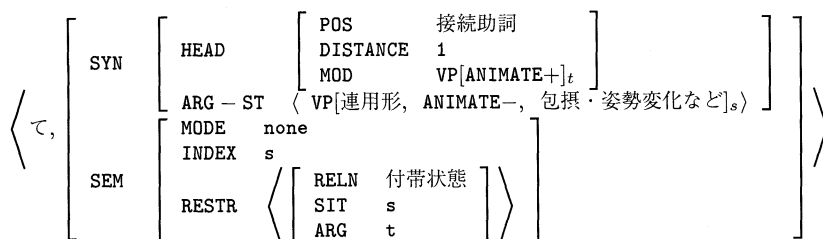
「疲れが出る」の主体が無生物、動詞が無意志動詞であるところから、「原因」を表す接続助詞「て」の接続関係パターンが適用されて、「原因」を表す連用修飾節として解析される。この連用修飾節は、係り先に、生物主体、無意志動詞を要求する。

(a) 接続助詞「て」による接続関係で「原因」を表す場合



注) CONJ[原因, DISTANCE 7, MOD[VOLITION-, ANIMATE+], ARG-ST[VOLITION-, ANIMATE-]]と略記する.

(b) 接続助詞「て」による接続関係で「付帯状態」を表す場合



注) CONJ[付帯状態, DISTANCE 1, MOD[ANIMATE+], ARG-ST[包摂・姿勢変化など, ANIMATE-]]と略記する.

図 6 辞書における接続関係パターンの記載例

4.2 接続構造の解析

連用修飾節の係り受け解析は、3章で述べたように、係り側の連用修飾節が要求する受け側の節の属性と、接続関係の关系的意味によって決まってくる連用修飾節の距離によって行う。

係り側の連用修飾節の距離が、受け側の連用修飾節の距離に等しいか小さいとき、すなわち、より密着しているときは、隣接する連用修飾節に係る。係り側の連用修飾節の距離が、大きいとき、すなわち、より離れているときは、飛び越えて先に係る。

この関係を、HPSG パーザに図 8 に示すような Head-Modifier ルールの制約として実装した。

〔例文 6〕にこのルールを適用して、接続構造の解析を行った結果を図 9 に示す。

〔例文 6〕 昼間の疲れが出て、杏子が母親の背中に負ぶさって、眠っていた。

連用修飾節「杏子が母親の背中に負ぶさって」は、動詞「負ぶさる」の意味分類が「包摂」を表すので、接続関係の关系的意味が「付帯状態」を表し、距離は 1 である。連用修飾節「昼間の疲れが出て」は、前述のように「原因」を表し、距離は 7 である。従って、「昼間の疲れが出て」は、「杏子が母親の背中に負ぶさって」を飛び越えて、主節の「眠っていた」に係る。連用修飾節の要求する主節の属性も、無意志動詞、生物主体であり一致する。主節の距離はもっとも大きく設定されているので、「杏子が母親の背中に負ぶさって」は、問題なく主節に係る。

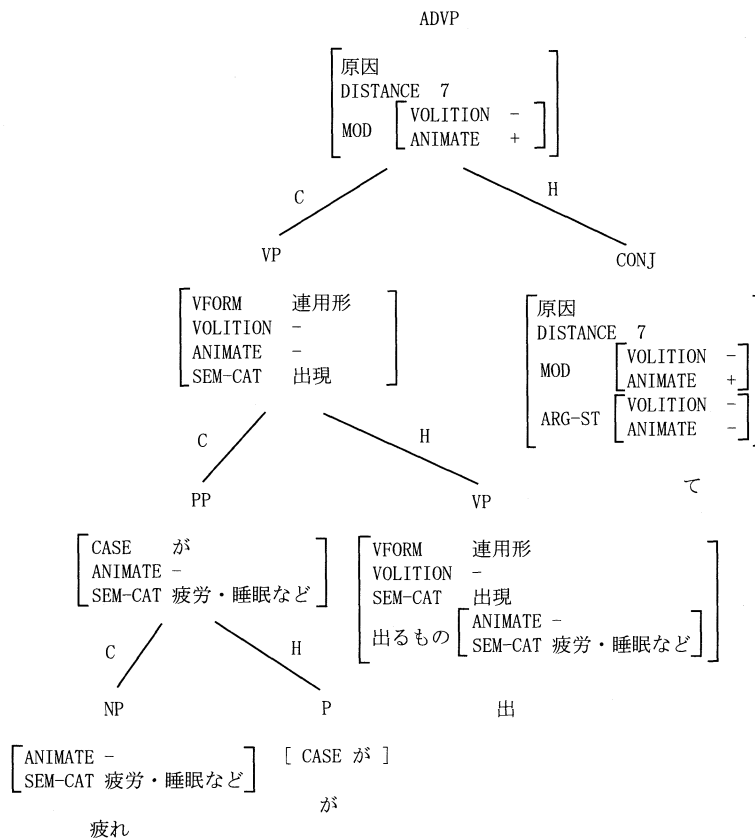


図 7 格パターンと接続関係パターンを用いた連用修飾節の解析

$$[\text{phrase}] \rightarrow \left[\begin{array}{l} \text{phrase} \\ \text{MOD } [1] \\ \text{DISTANCE } [2] \end{array} \right] H [1] \left[\begin{array}{l} \text{phrase} \\ \text{DISTANCE } [3] \end{array} \right] \{ \text{modifiable} [2, 3] \}$$

制約の定義

modifiable(1, 1)

modifiable(1, 2)

.

modifiable(8, 9)

図 8 Head-Modifier ルールに対する接続関係の距離による接続可能性の制約

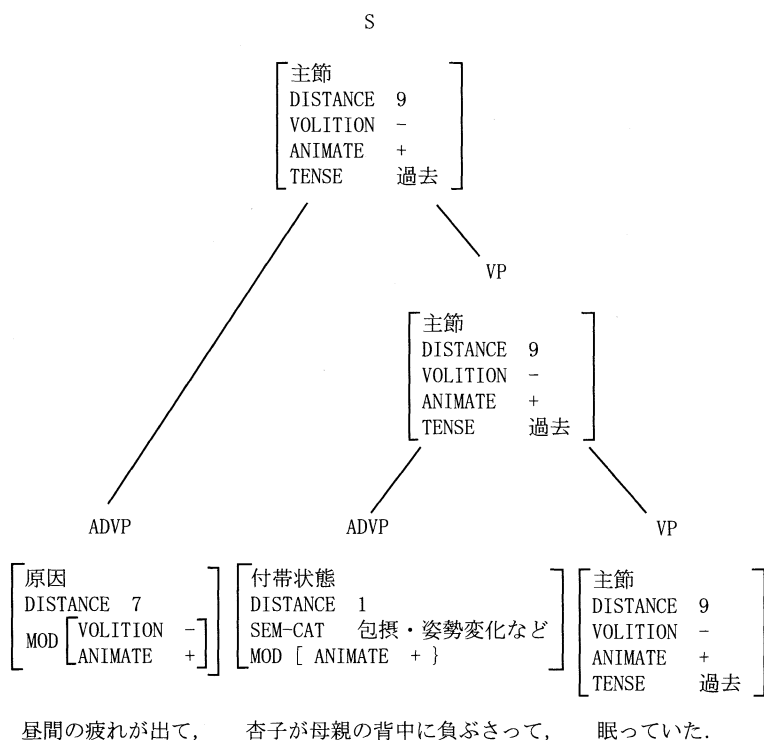


図 9 接続構造の解析

生成された意味構造を図 10 に示す。意味構造は「Syntactic Theory」(Sag and Wason 1999) によった。図はフラットな素性構造のリストで表されているが、INST(ANCE), SIT(UATION), ARG(UMENT) の変数により相互の関連を記述し、意味構造を表している。「原因」の連用修飾節 (SIT s) も、「付帯状態」の連用修飾節 (SIT u) も、「ARG t」と指定されており、共に主節 (SIT t) に係っていることが分る。

4.3 並列節の解析

並列節に対しては、一般の連用修飾節の係り受けとは別のルールが適用される。すなわち、並列の接続関係パターンに対しては、図 11 の Coordination ルールが適用される。係り側と受け側の節の動詞の意味分類、または各後置詞句を構成する名詞の意味分類のいずれかが一致するかどうかチェックされ、一致するときに並列節と解析される。Coordination ルールには、距離の制約条件がないので、並列の接続関係パターンは全ての節に適用可能である。並列節にさらに並列節に係ることも可能である。

〔例文 7〕のように、条件節に係った並列節を解析する場合の例を図 12 に示す。

〔例文 7〕 必要になったときにマウントして、不要になったときにアンマウントするので、……

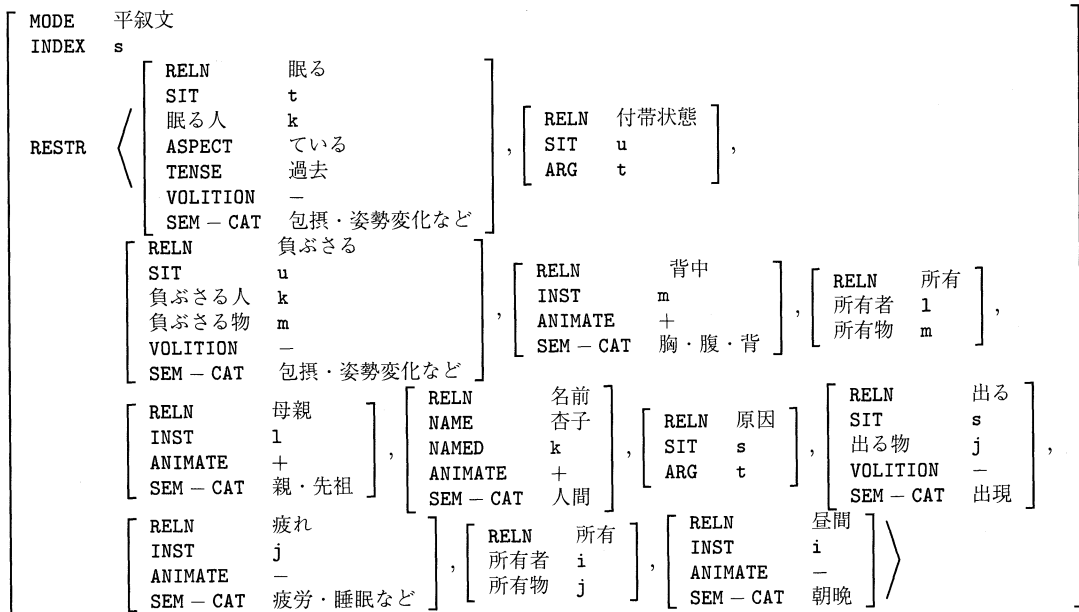
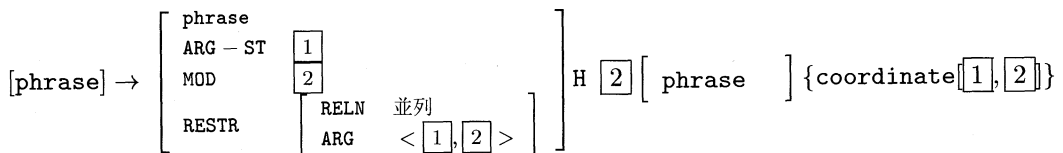


図 10 「昼間の疲れが出て、杏子が母親の背中に負ぶさって、眠っていた。」から生成された意味構造



coordinate[1, 2] : [1, 2] を構成する各節の対応する SEM-CAT に同じものがあるとき成立。

図 11 Coordination ルール

条件節と並列節の係り受けに対しては、通常の Head-Modifier ルールを適用する。この場合、並列節の距離は 8 に設定されているので、全ての節が係り得る。係り側が並列節のときは、Coordination ルールが適用され、条件節同士の意味分類が一致しているか、並列節と受け側の原因節の意味分類が一致しているかがチェックされる。条件節の動詞の意味分類が一致しており、並列節と原因節の動詞の意味分類が一致しているので、ルールが成立し、係り受けが成立する。この場合、原因節の距離が 7、並列節の距離が 8 であり、HeadModifier ルールでは係り受けが成立しないが、Coordination ルールでは、制約条件をチェックしないので、係り受けが成立する。

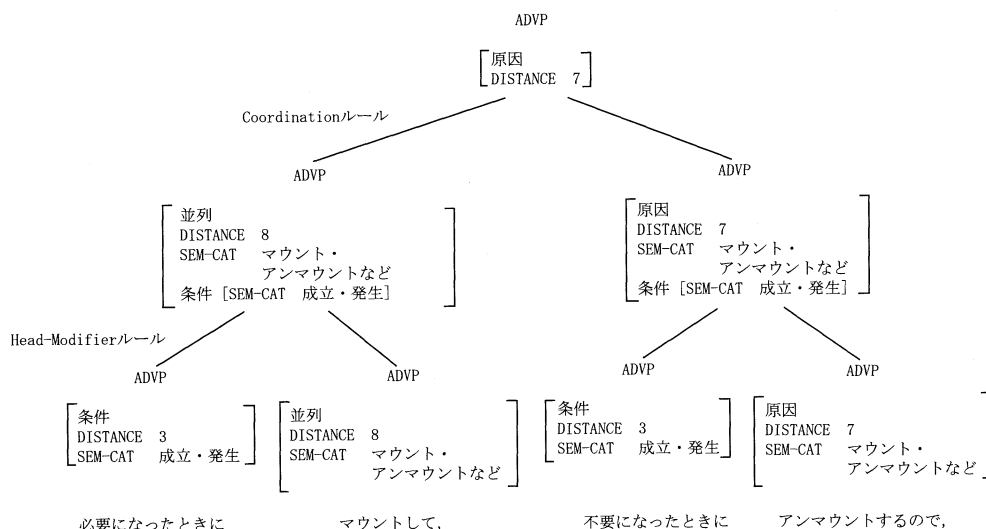


図 12 並列節の解析

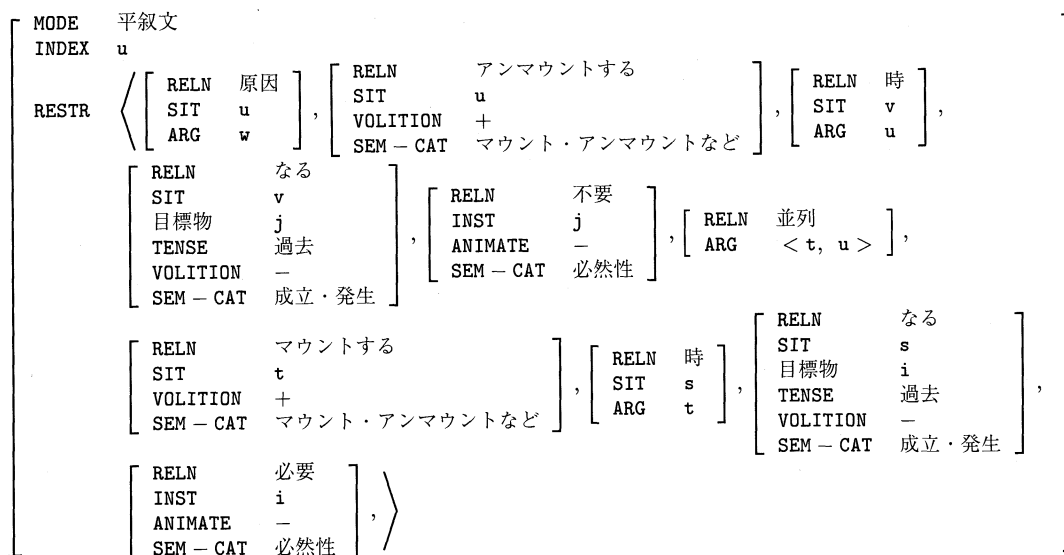


図 13 「必要になったときにマウントして、不要になったときにアンマウントするので、……」から生成された意味構造

生成された意味構造を図 13に示す。並列を構成する各々の節 (SIT t および SIT u) が、ARG<t, u>により並列を構成していることが分る。

5 接続構造解析モデルの評価結果

3章では、人手で解析を行ったが、同一の例文を、作成した接続構造解析モデルを用いて解析した。例文は、主節の外に連用従属節を2つ以上含む344文を用いた。

モデルから生成された解析結果の意味構造を分析した結果、3章で解析した結果より多少悪い90.7%の精度を得ることができた。これを、全体の2,010文に換算すると、98.4%に相当する。

間違った文を分析すると、並列の解析誤りが11%、接続関係の关系的意味の解析誤りが17%、その他72%がルール1, 2では正しく解析できない文であった。

接続の表現の頻度によるモデルも作成して解析した。解析結果は、3章で解析した結果とはほぼ同等の82.3%の精度を得ることができた。これを、全体の2,010文に換算すると97.0%に相当する。

6 むすび

本論文では連用修飾節の係り受けを解析し、接続構造を求めるために、接続関係の关系的意味を用いるモデルを作成し、実験した結果を述べた、接続の表現の曖昧性を解消して、接続関係の关系的意味を確立するために、動詞と主体の属性を用いて、接続関係をパターン化した。動詞の属性として、動詞の意志性、意味分類、慣用表現、ムード・アスペクト・ヴォイス、主体の属性として、主節と従属節の主体が同一かどうか、無生物主体かどうかを採用した。

本モデルを、実際の技術文書に適用して評価した結果、98.4%の正しい解析結果を得ることができた。关系的意味を用いないで、接続の表現の分類だけによった場合は97.0%の精度であったから本モデルの方法により誤り率が約半分に改善された。

参考文献

- 北条淳子 (1992). 談話の研究と教育, 7-112 頁: 複文文型. 国立国語研究所.
仁田義雄 (1995). 複文の研究 (上): シテ形接続をめぐる. くろしお出版.
情報処理振興事業協会 (1987). 計算機用日本語基本動詞辞書 IPAL(Basic Verbs). 情報処理振興事業協会技術センター.
金内典充, 今安正和 (1993). UNIX ネットワークプログラミング. オーム社.
国立国語研究所 (1989). 分類語彙表. 秀英出版.
南不二男 (1993). 現代日本語文法の輪郭. 大修館書店.
向仲 (1997). “動詞と主体の属性を用いた複文の接続関係の解析.” 自然言語処理, 4 (4), 3-16.
Sag, I. A. and Wason, T. (1999). *Syntactic Theory: A Formal Introduction*. CSLI Publication.

白井諭, 池原悟 (1995). “階層的認識構造に着目した日本語従属節間の係り受け解析の方法とその精度動詞と主体の属性を用いた複文の接続関係の解析.” 情報処理学会論文誌, **36** (10), 2353–2361.

宇津呂武仁, 西岡山滋之 (1999). “コーパスからの日本語従属節係り受け選好情報の抽出およびその評価.” 自然言語処理, **6** (7), 29–60.

略歴

向仲 景頁: 1953 九州大学工学部電気工学科卒業. 同年, 日本電気(株) 入社. 基本ソフトウェア開発に従事. 平成 5 年より金沢経済大学教授. 平成 9 年より江戸川大学教授, 現在に至る. 自然言語理解, エキスパートシステムの研究に従事. 情報処理学会, 言語処理学会, 人工知能学会, ACL, ACM, IEEE 各会員.

(2001 年 5 月 7 日 受付)

(2002 年 1 月 10 日 採録)